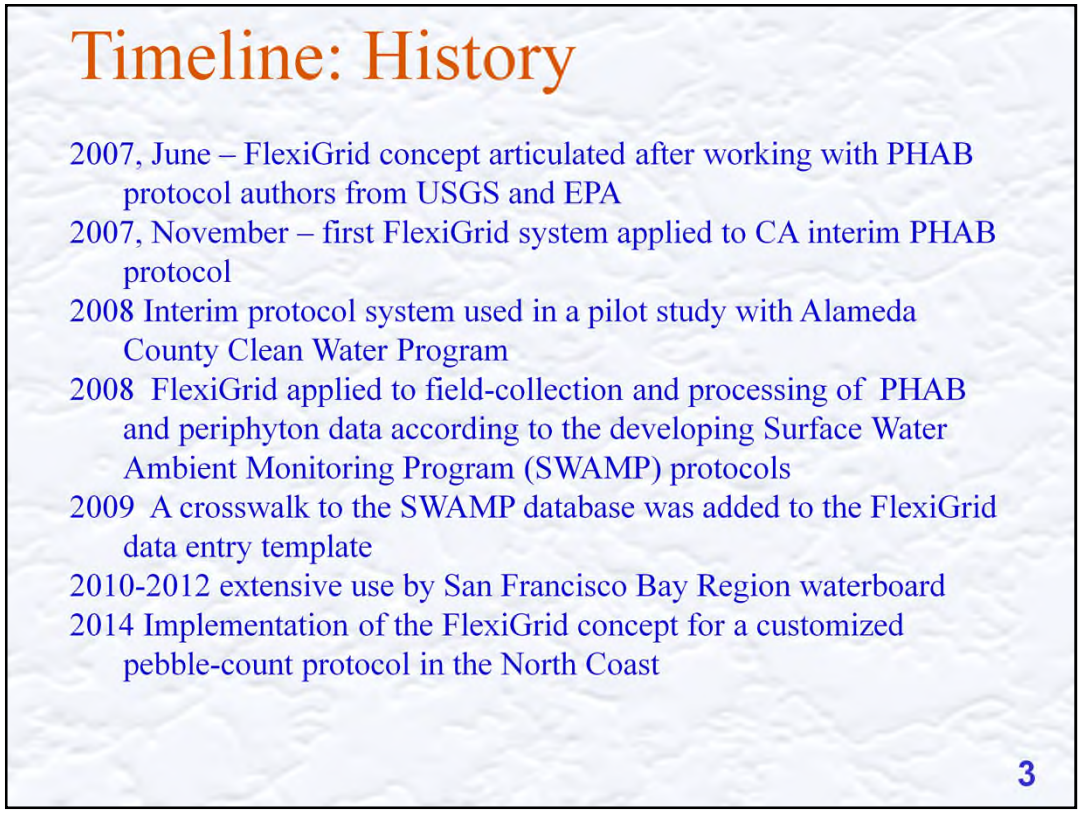
Managing Spatial Data: The FlexiGrid Experience

Revital Katznelson, Ph.D.
Berkeley, CA

Today's topics

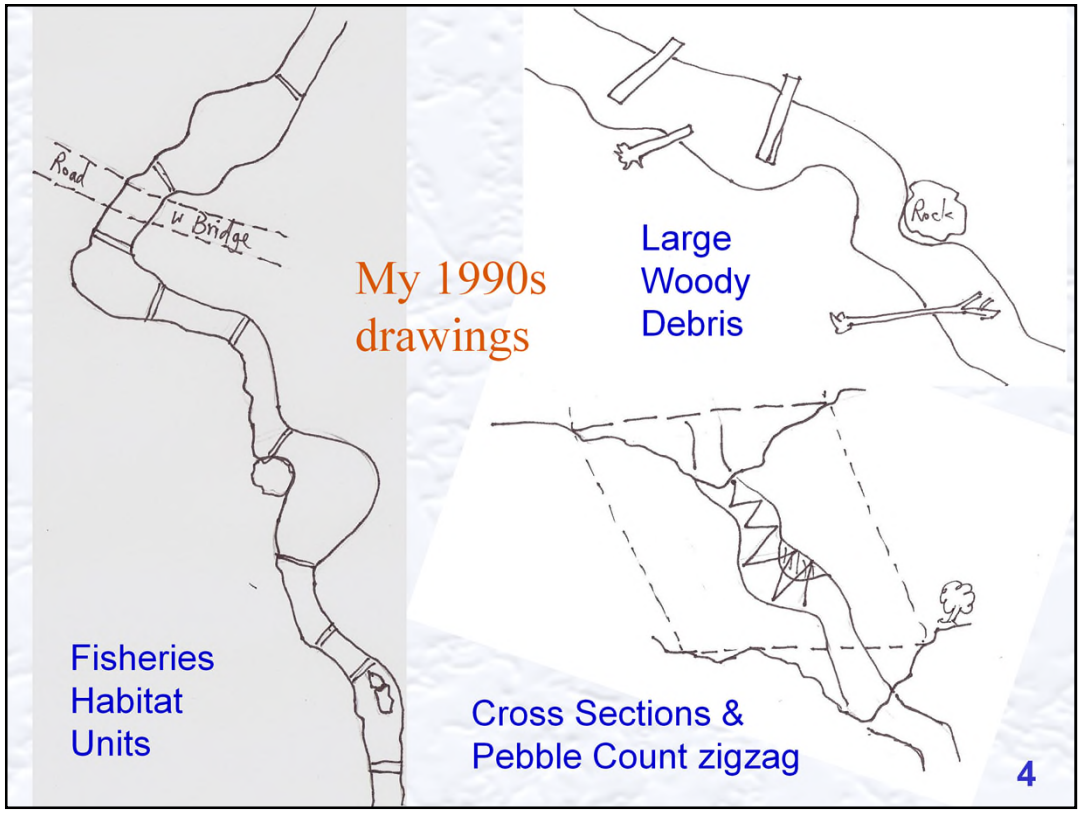
Timeline
Concept and Implementation
Lessons learned
Options

2



The idea was first depicted in a drawing and a spreadsheet in the summer of 2007. I created and implemented a functional data entry and Endpoint template system shortly after that, to accommodate an interim PHAB protocol; the customized FlexiGrid is still the only data processing option.

The system was practically re-designed to accommodate the new (2007) SWAMP protocol and the emerging capture of algae cover data; in one year we used 4 version of the SWAMP data sheets as they kept changing (talking about flexibility…). Some data were captured in the field using FlexiGrid template on a field computer. When the SWAMP database launched its PHAB module, I added a crosswalk for a seamless data transfer into the SWAMP database,



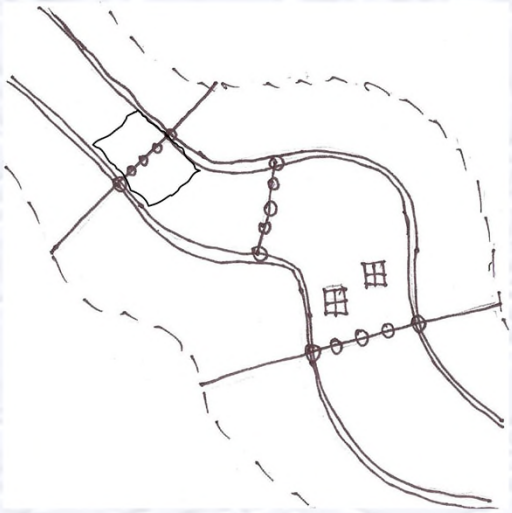
I have been working on different protocols that had spatial assessments for many years

Some of the protocols I encountered

- Fisheries habitat assessments DFG style (Flossi&Reynolds…)
- Thalweg profile & cross section surveys
- Pebble counts
- Large Woody Debris assessments
- Riparian plots & canopy density
- Residual pool volume (V*)
- Collection of vertical composite samples in water column
- Sampling storm runoff in outfall & creek networks
- Bird point-counts in a 50m-radius circle
- Flow discharge measurements
- Physical Habitat assessments RBP style
- Physical Habitat assessments EMAP style
- Physical Habitat assessments NAWQA style
- (more)

5

My 2000s drawings

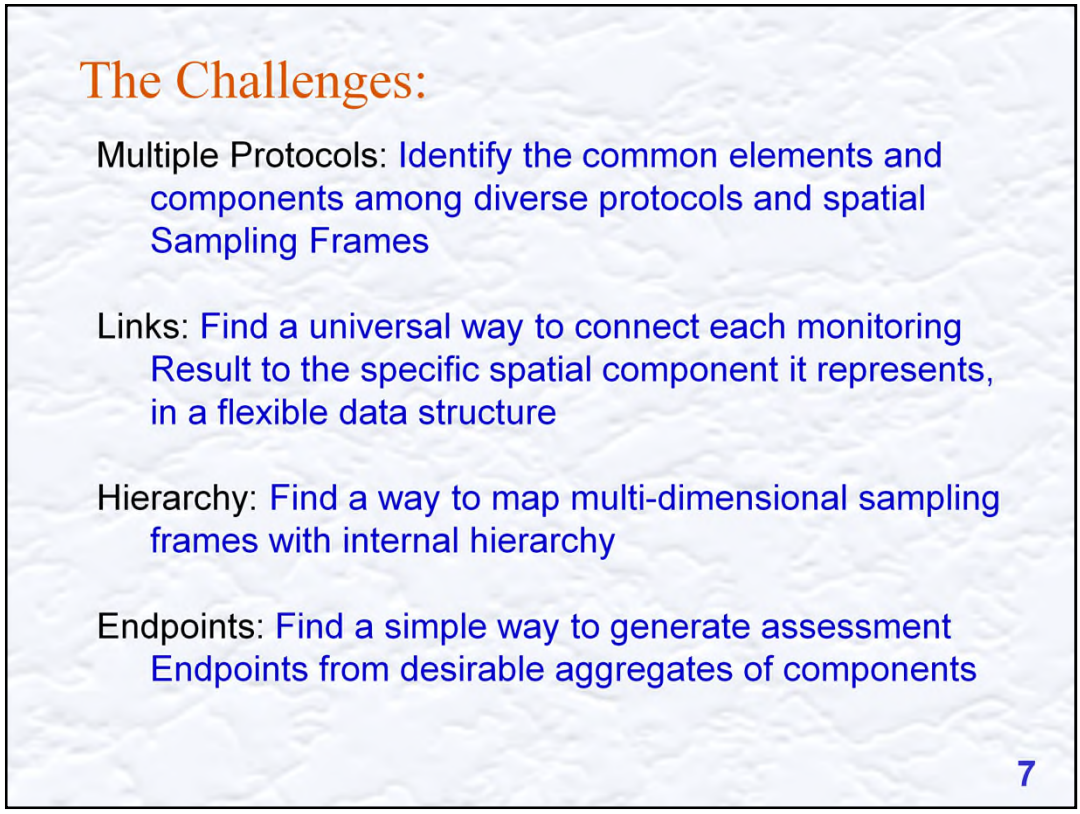


Transects
Transect points
Intertransects
Transect Plots
Riparian plots
Stream segments
…
…
…

Lots of “Crosslines”, and “Plots”
…and the internal hierarchy gets more and more complicated

6

When complex physical assessments became popular, spatial sampling frames became more and more elaborate and comprehensive



Problem: Spatial assessments are currently captured in data management systems that are very different from each other

Challenges

1. **Identify the common elements** shared by all spatial sampling frames, and devise 'neutral' terms for referring to them. This will help to increase data comparability by use of the same language. The proposed neutral terms are defined in the Glossary below.

2. **Create a data structure** that connects each monitoring Result (or Endpoint) to the exact spatial component (or aggregate thereof) in which it was measured, make the structure **flexible** to accommodate any protocol, and make it **compatible** with existing water quality data management systems. This will facilitate data exchange.

3. **Capture** the information about any spatial sampling frame or any of its components on a **virtual map** that can be stored in a database and/or easily transferred to a Geographic Information System (GIS). This will enable use of locative media to present and share results.

Spatial Sampling Frames

Can share the following properties:

- A 'Grid' made of a number of components
- Flexible (variable angles)
- Multi-dimensional
- A structure with internal hierarchy

They can easily be defined in neutral terms: each Frame has

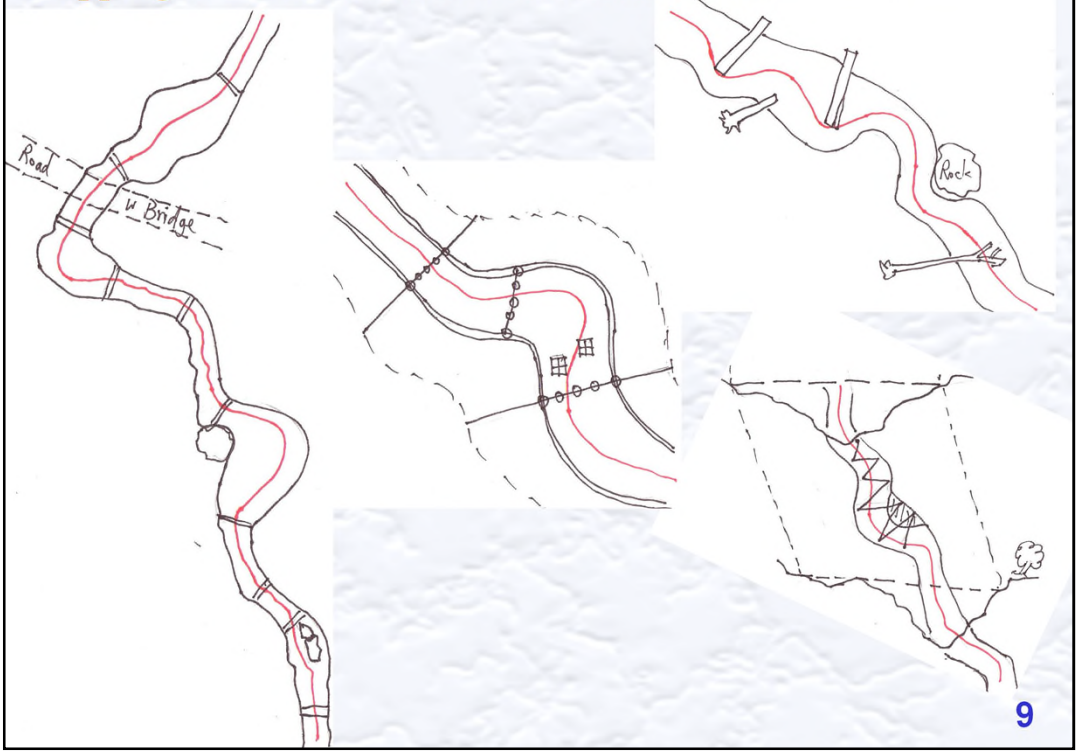
- An origin
- A direction (upstream or downstream)
- A known curvilinear distance.

The **Origin** can be mapped to a Permanent Station and/or Monument

Curvilinear distances can be measured on a tape placed on the thalweg, or the centerline, or the centroid…. Every Frame or grid has a **"backbone"** of sorts

8

Mapping is so much easier with a Backbone!



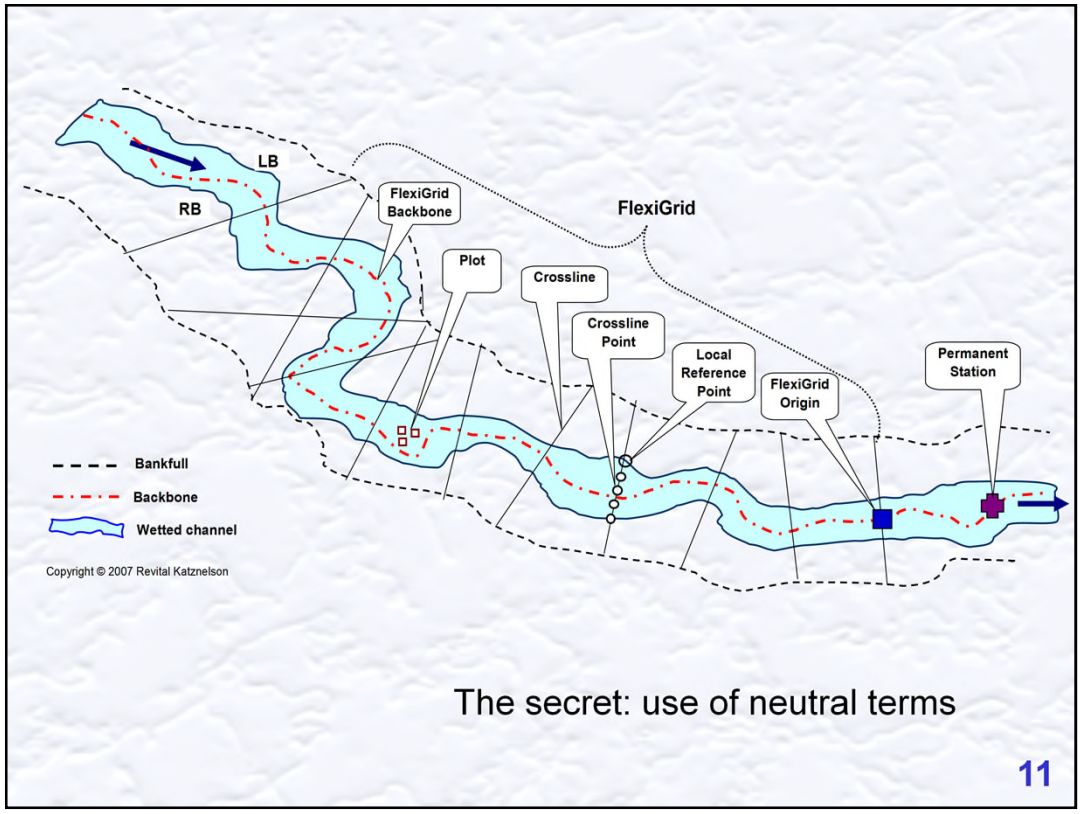
We are talking about “a spatial multi-dimensional flexible sampling frame with a variety of components, some organized with an internal hierarchy” (in short, SMDFAFWVCIH).

I called it “A FlexiGrid”

A FlexiGrid with a unique ID is laid out on a stream channel by one Team in one Station Visit. A Team can lay out one or more FlexiGrids in one Visit, each with its own unique ID.

The FlexiGrid refers to permanent locations, but is not permanent itself (FlexiGrids may be laid out differently at different times).

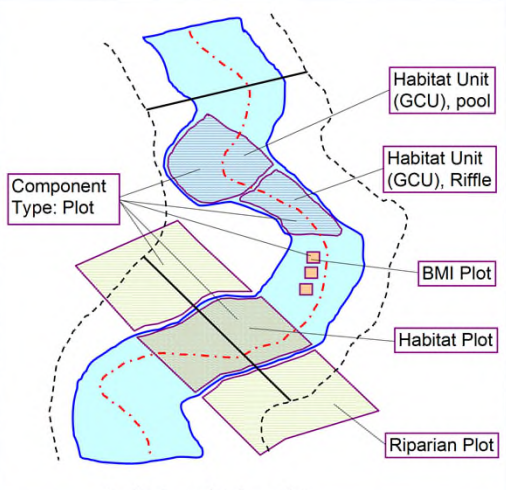
The FlexiGrid concept can be applied to different scales.



The secret: use of neutral terms

The FlexiGrid is a Concept

It is a way of organizing information about spatial components of complex sampling frames to enable:



A plot is a plot is a plot

(a) Description and mapping of the Frame and each component within the frame

(b) Linking each assessment Result to the component in which it was collected.

12

Common Data Elements

(These are added to the 'generic' Station-Visit descriptors:
Project, Trip, Team, Date, Time, Station, Station-Visit ID, etc.)

List 1: FlexiGrid (Frame) Descriptors

FlexiGrid **Origin ID**
FlexiGrid distance **units** (m, ft)
FlexiGrid Origin **distance from permanent Station** [/monument]
FlexiGrid **Type** (transects&points, right-angle grid, string of hab units, etc.)
FlexiGrid **Backbone** (Thalweg, midstream, centerline, centroid)
FlexiGrid Positive-distance **direction** (Upstream, downstream)

List 2: FlexiGrid component descriptors

FlexiGrid Component **ID** [naming conventions can reflect hierarchy]
FlexiGrid Component **Type** (e.g., Stream-fragment, Crossline, Plot, Point, Vertical, River-Parallel, Crossline-point, etc.)

FlexiGrid Component **Name in Protocol** (e.g., Reach, Transect, Transect-point, BMI Plot, etc.); [these are protocol-specific]

FlexiGrid Component Pre-determined **size [/shape, /boundaries]**
FlexiGrid Component **backbone-distance from Origin**
FlexiGrid Component place in series
FlexiGrid Component **distance from Local Reference Point**
Local Reference Point type (e.g., Left-bank)
FlexiGrid Component **depth from surface**

FlexiGrid and FlexiGrid component descriptors

Remember: Any of the spatial components you have defined can be included in the FlexiGrid; you do not HAVE TO snap them to the grid (i.e., to map them) with distances or depth measurements

But if you want to, you CAN!

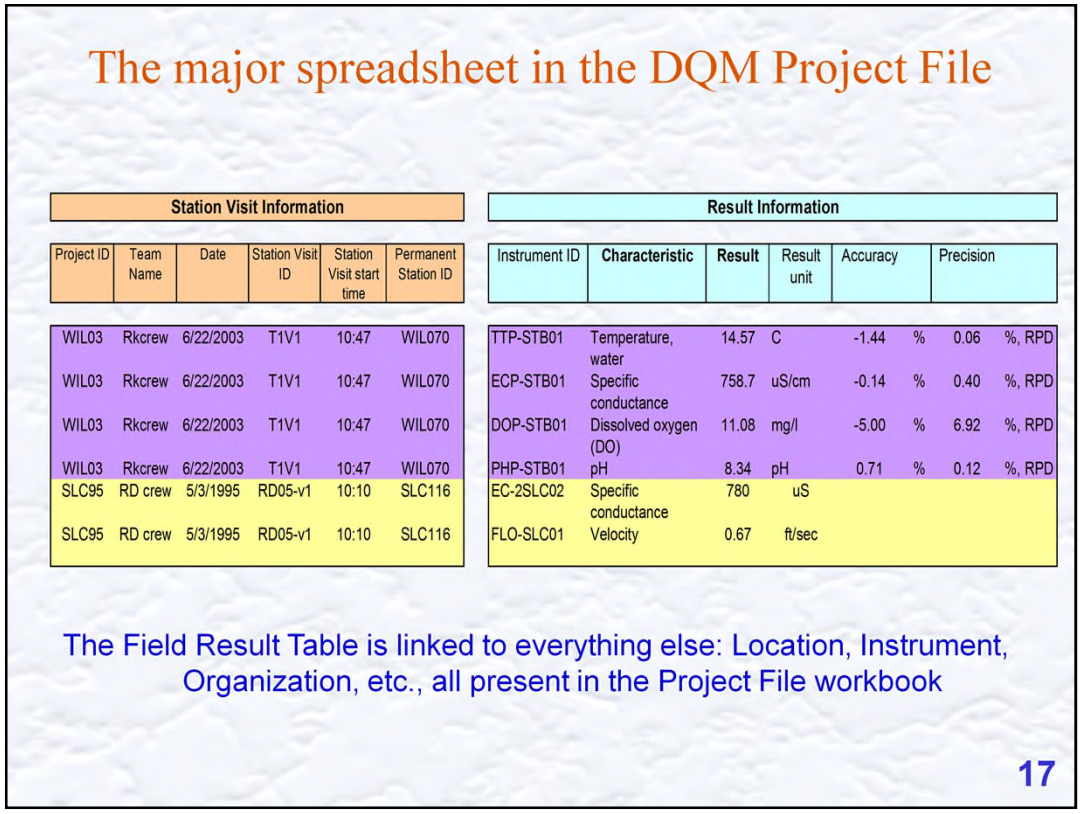
All you need is … a spreadsheet.

And I already had a spreadsheet. Actually, I had more than one, I had the entire workbook of the Project File.

Timeline: Pre-History

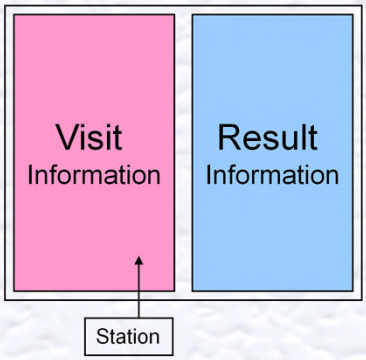
The Data Quality Management (DQM) system was developed for field measurements in 1998-2001, and implemented by the SWRCB Clean Water Team (Citizen Monitoring Program). In 2006 it was published as part of the Clean Water Team Toolbox and is available on line

http://www.waterboards.ca.gov/water_issues/programs/swamp/cwt_toolbox.shtml

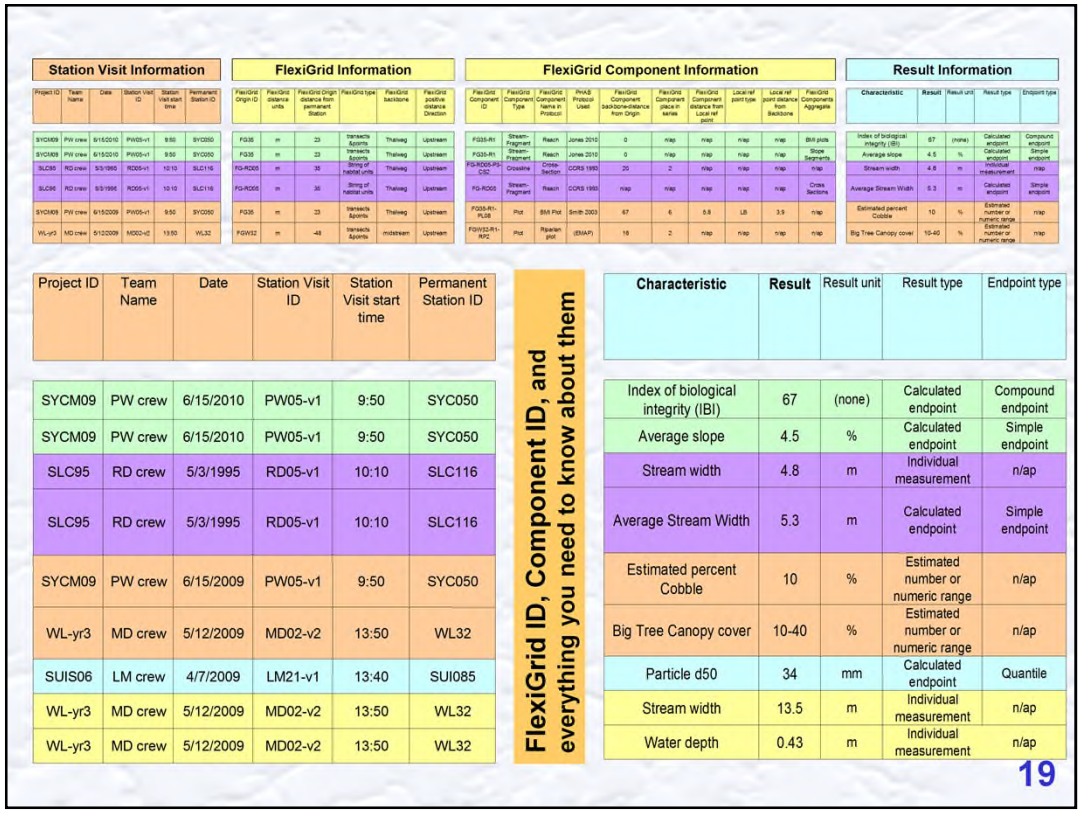


I developed a data quality management system built on Excel spreadsheets that capture all the bits of information one needs to see, but until 2007 it all focused on a point (or a line).

The basic Results spreadsheet: a schematic



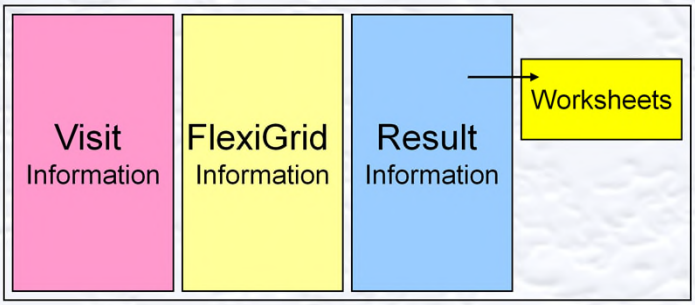
Everything was simple when the Station was a point.



Once you have identified the data elements, you can build a spreadsheet!

The FlexiGrid Results spreadsheet, Part 1

(Data Entry Template)



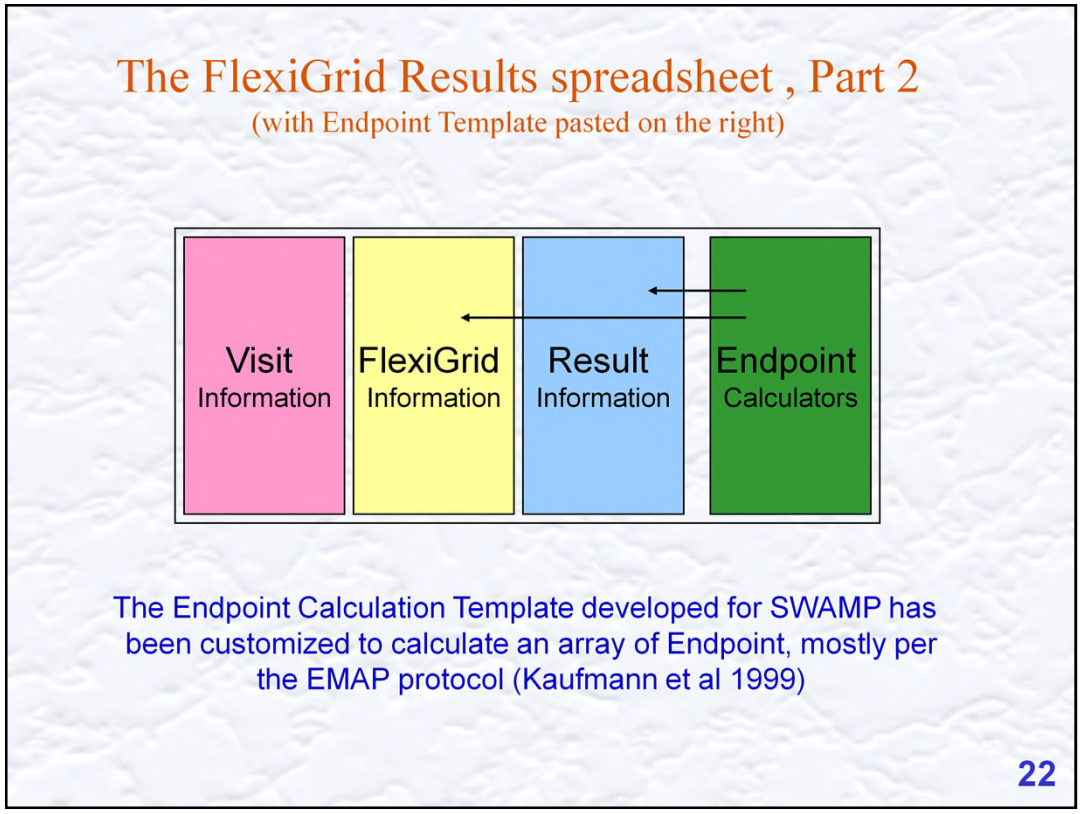
This spreadsheet has been customized several times.
Each version served as a **data entry template** for a specific SOP

Assessment Endpoints

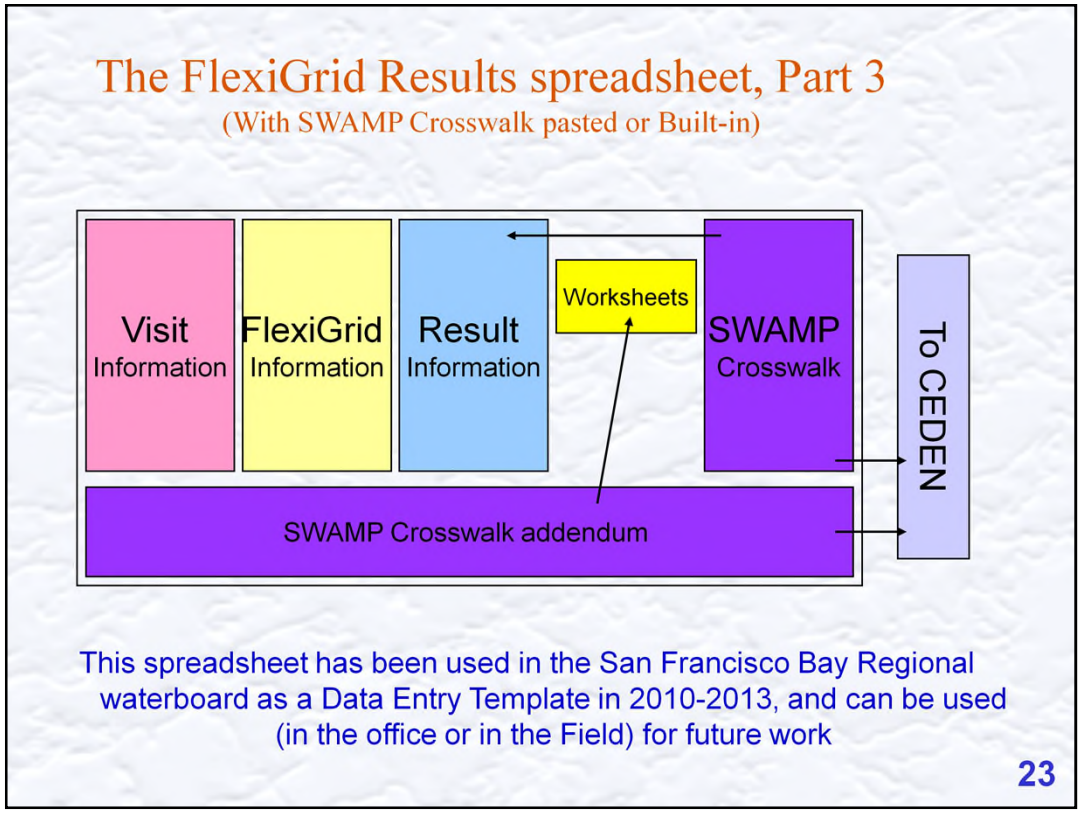
Descriptive statistics (Average, Frequency), Metrics, Indices, and any other Endpoint calculated for an aggregate of raw data.

Examples: Average wetted channel width, Percent cobble, Landfill/Trash Human Influence Index

Trip start date	SV start time	Permanent Station ID	FlexiGrid Component place in series	Characteristic	Result	Result unit	Processed Result (Formula!)		Result [Endpoint]	Unit	Characteristic [Endpoint]	Components Aggregate for Endpoint	Number of observations in aggregate (n)
4/13/2005	8:45	201EAS050	1	Boulders cover	1	(code)	5		11	%	Average Boulders cover	Habitat Plots	11
4/13/2005	8:45	201EAS050	2	Boulders cover	1	(code)	5						
4/13/2005	8:45	201EAS050	3	Boulders cover	0	(code)	0						
4/13/2005	8:45	201EAS050	4	Boulders cover	1	(code)	5						
4/13/2005	8:45	201EAS050	5	Boulders cover	0	(code)	0						
4/13/2005	8:45	201EAS050	6	Boulders cover	1	(code)	5						
4/13/2005	8:45	201EAS050	7	Boulders cover	2	(code)	25						
4/13/2005	8:45	201EAS050	8	Boulders cover	0	(code)	0						
4/13/2005	8:45	201EAS050	9	Boulders cover	2	(code)	25						
4/13/2005	8:45	201EAS050	10	Boulders cover	2	(code)	25						
4/13/2005	8:45	201EAS050	11	Boulders cover	2	(code)	25						



This system allowed for capturing, mapping, and calculating endpoints for several years of data when no alternative was available.



When the SWAMP raw PHAB data module and its upload template was functional, it was simple to add columns and rows to the data entry template (shown in purple); these purple cells are “looking” at red-yellow-blue cells in the original FlexiGrid spreadsheet – where the data is entered - and “translating” the contents into SWAMPese. The crosswalk products in the purple cells are easily isolated directly into the SWAMP upload template and transferred to SWAMP or CEDEN, the central data exchange network used in California.

Summary Part 1: The FlexiGrid concept supports a data structure that can do the following:

-- Connect each monitoring Result to the exact spatial component it was measured in (or to aggregate thereof);

-- Stage the raw data for easy, streamlined derivation of descriptive statistics, metrics, indices, and any other Endpoint;

-- Place every spatial component on a 'virtual map' as an independent entity while preserving the internal hierarchy and its order in a series;

-- Enable reconstruction of the Frame from the virtual map that is stored in database cells; and allow information transfer into a GIS system;

24

Summary Part 2: More good things FG can do

-- Integrate physical habitat (PHAB) results and descriptors with results and descriptors from any other area of inquiry (e.g., chemistry, toxicity, bioassessment, etc.);

-- Enable reporting of any type of Result and Endpoint, generated for any Characteristic in any spatial component and at any scale, on the 'same page'; and

-- Accommodate data generated by any Protocol that has multiple spatial components, including PHAB protocols used by different Agencies and Programs (e.g., NAWQA, EMAP, EPA RBP), DFG Fisheries habitat, etc.

Currently available: All FG templates for SWAMP, plus all SOPs for streamlined data processing and transfer

25

Summary Part 3: Drawbacks

-- Tailoring the FlexiGrid spreadsheet to a given protocol requires real understanding of that protocol

-- Development and customizing of the templates requires focused thinking and advanced Excel skills

-- Although data entry and streamlined endpoint derivation can be done by anyone, the Excel Data Entry Template cannot support as many 'fool-proof' safeguards against erroneous entries as some other database systems, and it is harder to protect

-- Users need training and technical support

26

The most practical drawback is that users need training and technical support.

Practically speaking …

-- Users need training and technical support

But this is true for any data management system. FlexiGrid help is available and it does not cost much.

The FlexiGrid is a good option for small monitoring entities that have no funding for massive data management and no access to IT support.

It is just an Excel spreadsheet!

27

Thanks!

San Francisco Bay Region (RB2) SWAMP Team 2007-2014:
Matt Cover, Karissa Anderson, Carrieann Lopez, Annika Anderson, Peter Otis, Karen Taberski, David Williams, Kevin Lunde, Caitlin White

Alameda County Clean Water program 2008
Arlene Feng, Paul Salop

North Coast Region (RB1) 2010-2014
Carrieann Lopez, Richard Fadness

Moss Landing Marine Lab SWAMP Data Management Team 2010-2014:
Marco Sigala, Cassandra Lamerdin